Materials and Methods

Materials and Methods are available in the online-only Data Supplement

Study Population

The CATHGEN cohort is a cardiovascular cohort of 9,334 adult patients recruited from the Duke University Cardiac Catheterization Clinic between January 2001 and December 2011.[1] Signed informed consent was obtained from all subjects prior to enrollment. Participants experiencing an acute coronary syndrome at the time of catheterization were excluded from enrollment as informed consent could not be obtained in these cases. The CATHGEN biorepository has been approved by and follows all Duke University Institutional Review Board policies.

The primary address of each CATHGEN participant as listed in their medical record was street-geocoded by the Children's Environmental Health Initiative.[2] Of the 9,334 individuals enrolled in CATHGEN, 86% (8,017) were successfully geocoded, and 25% (2,318) resided in Durham, Wake, or Orange counties, North Carolina (Figure 1). We restricted our analyses to residences of these three adjoining counties as the most complete traffic exposure data were available in this area.

CATHGEN is primarily composed of self-identified European-Americans (EA) and African-Americans (AA). For the purposes of this analysis, we removed those individuals who did not self-report as European-Americans or African-American (N = 64). We additionally removed individuals who resided more than two miles (3.22 km) from a primary or secondary roadway (N = 130). Individuals who resided more than two miles from a roadway may represent a very rural cohort whose exposure to traffic-related air pollution may not be well captured by distance to primary and secondary roadways, and may even require accounting for "traffic"

pollution from combines, tractors, and other industrial/agricultural sources. Removing these individuals left a study cohort of 2,124 individuals for this analysis (Table 1).

Clinical Information

To better understand associations across a spectrum of vascular diseases, we considered four cardiovascular outcomes: MI, the number of diseased coronary vessels at catheterization (DV), hypertension, and peripheral arterial disease (PAD). Hypertension was defined as an indication of a history of hypertension (determined either by patient report or a history of hypertension indicated in the medical record) on the health and physical exam administered by the attending physician or medical fellow as part of the cardiac catheterization. PAD was defined according to the Charlson Comorbidity Index,[3, 4] and also was taken from the aforementioned health and physical examination in the same manner as hypertension. MI was assessed as a recorded MI in the hospital records which occurred in the 7 days prior to the cardiac catheterization procedure. DV was a categorical outcome taking levels 0, 1, 2, and ≥ 3, and defined as the number of cardiac vessels with significant ($\geq 75\%$) blockage as assessed by a physician based on images from the cardiac catheterization.

We included sex, self-reported race, smoking, body mass index (BMI, kg/m^2), type 2 diabetes status, hyperlipidemia, and median home value at the census tract level as *a priori* confounders mirroring previous publications with this cohort.[2] Smoking was a binary variable with a positive indication defined as a current smoker (at least ½ pack per day), or a history of smoking within the previous 5 years that was terminated due to cardiac complications. This definition preserves those who have a history of smoking related cardiovascular complications. We did not include age as a covariate as it represents the age at catheterization and thus would be a covariate in the causal pathway (collider) as individuals with higher exposures may have

worse cardiovascular complications at younger ages. Adjusting for colliders such as this can introduce bias.[5, 6]

Traffic exposure metrics

We used the 2010 Topologically Integrated Geographic Encoding and Referencing North Carolina Primary and Secondary Roads State-based shapefile (4th quarter release) to define the roadway network within Durham, Wake, and Orange counties. Primary and secondary roads were defined in accordance with the Master Address File/Topologically Integrated Geographic Encoding and Referencing Feature Class Code definitions used by the United States Census, with primary roads (A1) defined as divided, limited-access highways, distinguishable by the presence of interchange, and secondary roads (A2) were defined as main inter- and intra-city arteries. We defined the distance to a roadway as the perpendicular distance from the primary residence to the nearest primary or secondary (major) roadway. We used a negative-logarithm transform of the distance to roadways as this might better approximate the decay in traffic-related air pollution as the distance to major roadways increases. We compared the use of this transform both to an untransformed distance to roadway exposure as well as two categorical assessments of traffic exposure. Except where explicitly stated otherwise, all associations with "distance to roadways" and "residential proximity to roadways" refer to those made when using the negative-logarithm transform.

We constructed six mutually exclusive (i.e., non-overlapping) traffic exposure zones (TEZs) which categorized exposure to traffic-related air pollution based on the following roadway and urban characteristics: high volume (>40,000 vehicles/day) traffic with regular congestion-related delays (TEZ 6); high volume (>40,000 vehicles/day) traffic with smooth flow (TEZ 5); mass transit routes (TEZ 4); urbanized areas with signal light density (TEZ 3); urbanized areas (TEZ 2); and the remainder of the study area (TEZ 1). [7, 8] A 1-km buffer around

the Raleigh-Durham airport was excluded from the study. TEZs 4, 5, and 6 were defined as 200 meter buffers around specific road segments using traffic model data supplied by the Institute for Transportation Research and Education of North Carolina State University. The boundaries for the urbanized areas (TEZ 2 and TEZ 3) were based on US Census classifications and signal light density (for TEZ 3) was calculated using kriging with a 200-meter grid. For each point on the map the point was assigned to the highest value TEZ containing that point (Figure 3). For example, in the event that a point overlapped with transit routes (TEZ 4) and high volume traffic with smooth flow (TEZ 5) it would be assigned to TEZ 5. TEZ 1 served as the baseline zone, and due to small numbers, TEZs 5 and 6 were combined (TEZ 5+) for all analyses. We also computed associations treating the TEZs as a linear exposure, under the assumption of equal changes in risk between successive TEZs. We refer to this test as TEZ trend or simply Trend. Both residential proximity to traffic and the TEZs correlate with measured traffic-related air pollution in our study area.[7-9]

Statistical Methods:

We estimated the association between our four cardiovascular outcomes and traffic exposure using R v3.3.0.[10] The risks of the binary outcomes (MI, hypertension, and PAD) were estimated using a logistic model with a quasi-binomial link function. Since the number of diseased coronary vessels was an ordinal variable we used a cumulative link model, as implemented in the R package ordinal[11] with a logistic link, which is equivalent to a standard proportional odds regression model. Associations were fit using a basic model adjusted for race and sex, and with a full model that adjusted for sex, race, body mass index, smoking, type 2 diabetes status, and median home value at the census tract level (2010 Census). In a sensitivity analysis we reintroduced the 130 CATHGEN participants excluded due to residing more than two miles from a roadway to the cohort and re-evaluated associations between proximity to

roadways and the four vascular outcomes. All results are given in terms of the odds ratio (OR) and 95% confidence interval.

For residential proximity to roadways, we examined race and sex-stratified models for all outcomes and examined if there was a significant interaction between race or sex and our exposure using a multiplicative interaction model. We did not examine stratified models for the TEZs because the small sample sizes can make such estimates unreliable. Associations were considered significant at $P < 0.05$.

To better allow for comparisons with previous results in the literature, we compared associations obtained using four means of classifying residential proximity to roadways: two continuous measures of residential exposure to traffic (the negative log transform of residential proximity to roadways and untransformed residential proximity to a roadways) and two categorical measures which are associated with some of the same vascular phenotypes examined here.[12-14] The first (Categorical 1000m) binned individuals into sets of 0-100 m, 100-200 m, 200-1000 m, and 1000+ m while the second (Categorical 200m) binned individuals into 0-50 m, 50-100 m, 100-200 m, and 200+ m. For both the continuous and categorical measures, we used the full adjustment model and for the categorical measures we included all bins in the model via a factor variable and used the bin representing the furthest residential distance to a roadway as the baseline. To compare models, we estimated the Akaike Information Criterion (AIC), the AIC weights (AICw), and the ratio of AICw (AICw ratio) for each model. The AICw ratio gives an estimate of how likely two models are relative to each other, with values above 1 indicating that the numerator model is the more likely of the two.[15] For all AICw ratio comparisons we used AICw from the model with the negative logarithm transformed distance to roadways as the numerator.

1. Kraus WE, Granger CB, Sketch MH, Donahue MP, Ginsburg GS, Hauser ER, Haynes C, Newby LK, Hurdle M, Dowdy ZE. A guide for a cardiovascular genomics biorepository: The cathgen experience. *Journal of cardiovascular translational research*. 2015;8:449-457
2. Ward-Caviness CK, Kraus WE, Blach C, Haynes CS, Dowdy E, Miranda ML, Devlin RB, Diaz-Sanchez D, Cascio WE, Mukerjee S. Association of roadway proximity with fasting plasma glucose and metabolic risk factors for cardiovascular disease in a cross-sectional study of cardiac catheterization patients. *Environmental health perspectives*. 2015;123:1007
3. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of chronic diseases*. 1987;40:373-383
4. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, Januel JM, Sundararajan V. Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173:676-682
5. Greenland S. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300-306
6. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *International journal of epidemiology*. 2010;39:417-420
7. Mukerjee S, Smith L, Brantley H, Stallings C, Neas L, Kimbrough S, Williams R. Comparison of modeled traffic exposure zones using on–road air pollution measurements. *Atmospheric Pollution Research*. 2015;6:82-87

8. Smith L, Mukerjee S, Kovalcik K, Sams E, Stallings C, Hudgens E, Scott J, Krantz T, Neas L. Near-road measurements for nitrogen dioxide and its association with traffic exposure zones. *Atmospheric Pollution Research*. 2015;6:1082-1086
9. Hagler GSW, Baldauf RW, Thoma ED, Long TR, Snow RF, Kinsey JS, Oudejans L, Gullett BK. Ultrafine particles near a major roadway in raleigh, north carolina: Downwind attenuation and correlation with traffic-related pollutants. *Atmospheric Environment*. 2009;43:1229-1234
10. Team RC. R: A language and environment for statistical computing. 2009
11. Christensen RHB. Ordinal—regression models for ordinal data. *R package version*. 2010;22
12. Kirwa K, Eliot MN, Wang Y, Adams MA, Morgan CG, Kerr J, Norman GJ, Eaton CB, Allison MA, Wellenius GA. Residential proximity to major roadways and prevalent hypertension among postmenopausal women: Results from the women's health initiative san diego cohort. *Journal of the American Heart Association*. 2014;3
13. Hoffmann B, Moebus S, Möhlenkamp S, Stang A, Lehmann N, Dragano N, Schmermund A, Memmesheimer M, Mann K, Erbel R. Residential exposure to traffic is associated with coronary atherosclerosis. *Circulation*. 2007;116:489-496
14. Hoffmann B, Moebus S, Dragano N, Mohlenkamp S, Memmesheimer M, Erbel R, Jockel KH. Residential traffic exposure and coronary heart disease: Results from the heinz nixdorf recall study. *Biomarkers*. 2009;14 Suppl 1:74-78
15. Wagenmakers E-J, Farrell S. Aic model selection using akaike weights. *Psychonomic bulletin & review.* 2004;11:192-196